Exponent®

EXTERNAL MEMORANDUM

TO: Scientific Advisory Board (SAB) Members
Genevieve Matanoski, MD, Dr.PH, Chair

Thomas Miller
Environmental Protection Agency

FROM: Pamela Mink, Ph.D., M.P.H.

DATE: January 17, 2006

PROJECT: BE02909.001 0401

SUBJECT: Comments on the EPA SAB Draft Report, "Advisory on EPA's Assessments of Carcinogenic Effects of Organic and Inorganic Arsenic: An Advisory Report of the US EPA Science Advisory Board"

This memo summarizes my comments on the draft report of the EPA Science Advisory Board, Arsenic Review Panel (December 27, 2005). My comments relate to inorganic arsenic, and focus on the following issues:

1. Choice of additional databases for use in risk assessment
2. Choice of epidemiologic studies of populations exposed to low levels of inorganic arsenic in drinking water to be used in evaluating concordance with estimates from extrapolation of the Taiwan data
3. Use of an "integrative analysis" of exposure-response models
4. Brief comments on methodologic issues
    a. Bias due to misclassification
    b. Statistical power

1. Choice of additional databases for use in risk assessment and dose-response modeling

The SAB recommends that all of the relevant studies of arsenic in drinking water, including studies of both "high" and "low" exposure levels, be reviewed critically using a uniform set of criteria (p. 25 SAB draft report). The SAB also recommended that this assessment be presented in a tabular format that would include relative strengths and weaknesses, study caveats and assumptions, and variability within and among the studies. This will be helpful to the reader and is a great improvement over previous EPA and NAS documents, where the consistent use of uniform criteria to assess each study was not apparent.

With this in mind, it seems premature to determine that the studies by Ferrreccio et al. (2000) and Smith et al. (1998) conducted in Chile had "excellence of exposure assessment" as one of their strengths (p. 24, SAB draft report). This conclusion should await the full review of the studies based on the set criteria. In addition, before these studies are used to "scale the unit risks at high exposure levels that emerge from the Taiwan data," as proposed in the draft report (p. 24), limitations of the study that may impact interpretation of study results must be considered and weighed. Indeed, while previous EPA and NRC reports have included exposure assessment among the strengths of the Ferreccio et al. (2000) study, the EPA and NRC reports and the authors themselves have also pointed out that the control selection methodology was a major limitation of the study. Furthermore, the authors indicate that the direction of bias may differ for low (upward bias in observed relative risk) and high (downward bias in observed relative risk) exposure levels. Even if exposure assessment is excellent in this study, there could be considerable bias in the results due to issues with control selection. Thus, the decision as to whether these or other studies (e.g., Chiou et al. 2001; Chen et al. 2004) are likely to be useful to "scale the unit risks" from the Taiwan data should be made after a careful and thorough evaluation has been completed according to the recommendations of the SAB (p. 25, SAB draft report). After such a review, if it is determined that it is appropriate to conduct a risk assessment using one or more of these (or other) databases, any caveats and/or potential limitations should be described in the interpretation of the results.

With respect to charge question D-2 regarding the use of a linear model, it is also premature to recommend using data from Taiwan, Chile and Argentina, while dismissing studies by Lamm et al. (2004), Bates et al. (2003), and Steinmaus et al. (2003) (see p. 33, SAB draft report), until a full evaluation of the relative strengths and limitations with respect to the set of specified criteria has been completed. In addition, although the draft report indicates that studies from Chile and Argentina "seem to support a possible linear dose-response between exposure from drinking water and internal cancer risks" (p. 33 SAB draft report), a formal evaluation of linearity has not been presented. Therefore, decisions about which studies to use (and how they should be used) should be made after appropriate and formal evaluations of the studies have been completed. As discussed in the draft, this process should be transparent (see p. 25, SAB draft report), and the results of proposed sensitivity analyses and rationale for decisions regarding which data are used for the purposes of dose-response analyses should be made publicly available.

2. Choice of epidemiologic studies of populations exposed to low levels of inorganic arsenic in drinking water to be used in evaluating concordance with estimates from extrapolation of the Taiwan data.

The SAB's draft report lists the studies from the US and other populations with low-level exposure to arsenic in drinking water (i.e., 0.5 to 160 μg/L inorganic arsenic in drinking water) that the Panel suggests should be considered for use in assessing concordance with risk estimates derived from the Taiwan data (p. 24, SAB draft report) and that these studies should also be evaluated in sensitivity analyses to estimate the potential impact of systematic error (pp. 25-26). The studies listed in the report are not consistent, however, nor do they include all of the studies that appear to meet their criteria for exposure range and similarity to the US population. Both lists (p. 24 and p. 26) include the following studies:

- Bates et al. 1995
- Lewis et al. 1999
- Steinmaus et al. 2003
- Bates et al. 2004

The list on p. 24 includes Michaud et al. 2004, whereas the list on p. 26 includes Kurttio et al. 1999. These two studies should be on both lists. In addition, we recommend that the study by Karagas et al. (2004), conducted in New Hampshire, be added to the list. Finally, the study by Chiou et al. (2001) from Northeastern Taiwan includes adjusted relative risk estimates for the following categories of exposure: 0-10 μg/L, 10.1-50 μg/L, 50.1-100 μg/L, and greater than 100 μg/L. Thus, this study should also be considered in the assessment of concordance with particular dose-response models derived from populations with higher arsenic levels in drinking water.

As noted in Section 1 (above) and in the SAB draft report (p. 25), the selection of studies to be used in assessing concordance should be based on a priori review and inclusion/exclusion criteria, and the application of these criteria as well as the basis for the final set of studies should be transparent.

3. Use of an "integrative analysis."

We agree with the Panel's recommendation of an integrative analysis to assess concordance with exposure-response models derived from the primary analysis (p. 27, SAB draft report). An

integrative analysis could also be used to derive a slope factor, which would incorporate data from more than one study. As described in the SAB draft report (p. 26, SAB draft report), this approach has been used do estimate the association between methylmercury exposure and IQ. In the case of methylmercury, data on cognitive outcomes in addition to IQ were included in analyses of three cohort studies in order to derive a regression coefficient to represent the association between exposure (methylmercury) and outcome (IQ). A similar approach could be evaluated for use in estimating (i.e., what is the magnitude?) and characterizing (i.e., linear vs. non-linear) association between inorganic arsenic in drinking water and cancer.

Thus, we propose that the SAB recommends that the EPA conduct integrative analyses for two purposes:

1) To assess concordance with the exposure-response models:

   - NOTE: This approach was used in the meta-analysis conducted by Exponent. Specifically, we derived a summary relative risk estimate based on the studies of "low level" exposure to arsenic in drinking water to determine whether there was concordance between the relative risk of bladder cancer in these populations and the range of relative risks predicted by the models described in the 2001 NRC report. We found that for non-smokers, the meta-relative risk from these studies was below the range predicted by the dose-response models based on SW Taiwan data as presented in the NAS report (2001).

   - Criteria and rationale for including and excluding studies from this analysis should be publicly available and the process should be transparent.

   - Given that data from individual epidemiologic studies, results of the meta-analysis conducted by Exponent, and dose-response modeling presented by Dr. Kenneth Brown indicate a lack of concordance with current exposure-response model derived from the Taiwan data, options for the "next steps" should be clearly delineated. In other words, if the results of the "integrative analysis" suggest a lack of concordance, it is not clear how EPA will proceed in order to "correct" the exposure-response model? Thus, guidelines for potential approaches will be useful.

   - Results of this analysis should be made publicly available including a full description of the analytic procedures, and the public should be given the opportunity to read and comment on these results.

2) To derive a slope factor that is predictive of cancer risks across the spectrum of exposure levels:

- This "step" would involve conducting analyses and evaluating models that are based on studies in addition to the studies conducted on the SW Taiwan population.

- The analyses cited in the SAB draft report (p. 26) regarding methylmercury exposure and IQ can be consulted as an example of how this approach has been used to estimate slopes (regression coefficients), as well as related sensitivity analyses.

The full approach to conducting integrative analyses and how this approach could be applied (e.g., beyond assessing concordance) needs to be elucidated, transparent, and made available to the public for review and comments. We agree with the recommendation in the SAB draft report to conduct an integrative analysis and feel this approach can be useful not just in evaluating concordance, but also in deriving a slope factor and developing dose-response models. We encourage the SAB and EPA to consider application of this approach to the derivation of a slope factor and as a basis for dose-response modeling.

4. Brief comments on methodologic issues.

a. Bias due to misclassification

We agree with the Panel's recommendation to formally evaluate and quantify potential bias in the epidemiologic studies (p. 25, SAB draft report), including the potential effect of bias due to exposure misclassification.

The Draft SAB Report states, "Misclassification of exposure in such studies (when non-differential) can have a profound effect in depressing the magnitude of the observed risk." Nevertheless, relative risk estimates for virtually all categories of exposure in the "low level" studies are below 1.0 for non-smokers. In such cases, if there is non-differential misclassification and bias in the direction of the null, then the "true" relative risk would be even more extreme, giving the appearance of a protective association. The scenario that appears to be assumed in the statement quoted above is that the observed relative risk estimate is slightly above 1.0, so the "true" relative risk must be of (much) greater magnitude. When considering the potential role of non-differential misclassification, it is imperative to first evaluate whether the observed relative risk estimate is above or below 1.0. Thus, bias in the direction of the null could be either downward toward 1.0 (if the observed RR is greater than 1.0), or upward toward 1.0 (if the observed RR is less than 1.0).

b. Statistical power

The SAB draft report refers to four studies that it characterizes as "underpowered" (Lamm et al., 2004; Bates et al., 2003 (*sic*); Steinmaus et al., 2003) or as having "limited power" (Chen et al., 2004) (p. 33, SAB draft report). Again, a formal evaluation of statistical power should be part of the evaluation of the epidemiologic studies, and results and conclusions should be presented as part of this transparent process. For example, it is not clear why the SAB draft report states, "…this study also has limited power to examine the form of the dose-response relationship within the 10-100 μg/L range (Chen et al 2004)" (p. 33, SAB draft report) given that over half of the person-years in the study occurred in this exposure range. Furthermore, the sample size and power calculations for the Steinmaus et al. (2003) study in the EPA Toxicological Review are incorrect and grossly overestimate the sample size required to detect a relative risk of approximately 1.2. Finally, the study by Bates et al. (2004) was not evaluated by the NRC or EPA documents and the basis for the SAB Panel's conclusion that it is "underpowered" is not presented.

As stated in my previous comments to this SAB (October 14, 2005), we agree with the limitations and cautions presented by Checkoway et al. (2004) with respect to statistical power and sample size. Based on this, statistical power should not be used as the sole basis to include or exclude a given study from consideration. Furthermore, studies that may not have sufficient power to detect a modest relative risk can still be combined with other appropriate studies in a meta-analysis or other integrative analysis to increase the statistical power and to improve the precision of the overall (combined or meta-) risk estimates.